Purdue University
Purdue e-Pubs

Department of Computer Science Technical Reports
Department of Computer Science

1994

On The Three Diumensionality of Motion Correspondence

Anupam Joshi

Zygmunt Pizlo

Report Number:
94-007

ON THE THREE DIMENSIONALITY OF MOTION CORRESPONDENCE

Anupam Joshi
Zygmunt Pizlo

CSD-TR-94-007
January 1994
(Revised February 1994)

On the three dimensionality of motion correspondence

Anupam Joshi
Department of Computer Sciences
Purdue University
West Lafayette, IN 47907 USA

Zygmunt Pizlo
Department of Psychological Sciences
Purdue University
West Lafayette, IN 47907 USA

Abstract

Motion correspondence has long been regarded by psychophysicists as being performed with respect to two dimensional distances only. This is despite the fact that obtaining correspondence of 3D objects from 2D positions is a mathematically ill posed problem. In this work, we reinterpret extant psychophysical experiments that seemingly support the claim about two dimensionality of correspondence, and discuss available evidence from neuroanatomy and psychophysics which seems to indicate that correspondence problem is solved by a three dimensional process. Brief remarks about the significance of this analysis for computer vision are also made.

Keywords Correspondence Motion

Motion correspondence is defined by Ullman [31] as the process by which elements in different views are identified as representing the same object at different times, thereby maintaining the perceptual identity of objects in motion (Object Constancy). It can be said *sans* hesitation that the problem of obtaining correspondence is a fundamental aspect of computational vision and underlies much work on motion. It has been traditional to discuss motion correspondence with respect to two dimensional images of the three dimensional world [25, 6, 32, 1, 10, 23, 33, 11]. This, however, poses a problem because the projection scheme involves a mapping which is many to one. As a result, the problem of image interpretation is ill posed. In other words, solving correspondence in two dimensions can lead to erroneous solutions, because points that seem to correspond in two dimensions may not be corresponding in three dimensions. The obvious solution would be to solve the correspondence problem in three dimensions. At this point, the reader may well be prompted to ask, "So why is most of the work in correspondence done with two dimensional data ?" The question is relatively easy to answer from the perspective of computer vision research, and the answer is simply that till recently, obtaining three dimensional (or range) data about the position of feature points was fairly difficult. Laser range sensors were few and far between, whereas camera like devices which produce two dimensional images were quite common in computer vision applications. In principle, depth maps could be produced from two 2D perspective images by using one or more depth cues (shading, texture, binocular disparity etc.) that are present in 2D images [13]. However, there are still no algorithms that can always produce reliable reconstructions of 3D scenes from 2D images. On the other hand, range sensing devices are becoming more and more accessible, and range sensors which can operate at video frame rates are now under development [5]. Thus purely from the computer vision point of view, it now makes sense to try and obtain algorithms which work on range data and use 3D correspondence.

Consider now human vision. We first discuss results of psychophysical experiments and then results of anatomical and physiological studies of the brain. This analysis will lead us to the suggestion that depth is involved in motion correspondence, which means that the correspondence is established in 3D. Our suggestion is based amongst others on the observation that ordinal, rather than metric, properties of depth are likely to affect motion correspondence because the former, not the latter, can be reliably perceived by the observer.

The primary source of the assertion, commonly accepted by psychophysicists, that correspondence is performed in two dimensions are some interesting, but not free from flaws, experiments done by Ullman [30]. Ullman presented subjects with stimuli consisting of perspective drawings of three dimensional shapes. All displays contained one line (C) shown in alternation with two others, one to its right (R) and one to its left (L). The displays were such that the distances L-C & R-C could have different values depending whether they are measured on screen, or in the simulated 3D space. In these experiments, Ullman consistently obtained the result that C was corresponded by the human subject with R or L depending on which was closer in terms of the 2D distance on the screen, even if the distance in the simulated 3D space was actually shorter for the other. This seemingly implied that correspondence is done in two dimensions and that depth is not involved at all. What has been overlooked, however, is the fact that these experiments were conducted with perspective images of 3D objects shown on a screen, and not with actual 3D objects. Binocular depth cues and accommodation were thus removed, and the three dimensionality of the object came solely from the perspective drawing being perceived as such. It is perhaps not very surprising that under such conditions, where 3D cues were greatly removed, the correspondence was done with respect to two dimensional distances. A similar observation was made by Mutch, Smith and Yonas [17], who sought to verify Ullman's results by performing a similar experiment, but this time with real 3D stimuli and with many depth cues available. Interestingly, they obtained results analogous to Ullman's. They too found that motion correspondence was determined primarily by 2D distances and depth did not seem to be involved at all.

It has to be pointed out, however, that the conclusions provided by Ullman and by Mutch *et al.* involve a rather strong implicit assumption. They assumed that the subject can reliably compare 2D and 3D distances, which is equivalent to the claim that the visual space is metric. This, however, does not seem to be true. It was shown that the human observer cannot reliably and accurately judge 3D distances[29, 28, 27, 16, 4, 21, 24]. This fact is consistent with the measurements reported by Mutch *et al.* A perusal of their data shows that their subjects greatly underestimated distances in 3D. All these results mean that the subjects do not have perceptual access to metric properties of the physical space and, therefore, the primacy of the 2D distances in Ullman's and Mutch *et al.*'s motion correspondence experiments tells little, if anything, about the possible role of depth in motion correspondence. Such a role can be adequately tested only if the stimuli use some aspect of 3D distance (or depth in particular), which is reliably perceived by subjects. It was shown by Todd and colleagues[27, 28, 29] that the observer has reliable access to affine properties, specifically, to the order in depth. As such, stimuli using this feature would provide a good test for any hypothesis regarding the 2D/3D nature of the correspondence process. If the order of the targets in depth does not affect the motion correspondence, the claim about the primacy of 2D distances on the image plane would be strongly supported. Interestingly, an experiment based on the ordinal aspect of depth has been performed by Ramachandran and Anstis [22] (they, however, did not discuss their result in the context of the controversy about 2D/3D mechanism of motion correspondence).

In this experiment, subjects were presented with an apparent motion sequence. In the first case, two vertically aligned spots are presented in the first frame, followed by the two spots shifted horizontally for the second frame. All four spots were coplanar, and in the frontoparallel plane. It was observed, consistent with earlier reports, that the apparent motion perceived was always horizontal, and a diagonal motion, which would lead to a crossing paths that were longer, was never observed(Figure 1). In the second case, two diagonally opposite spots were placed in a different stereoscopic plane from the other two (leading to non coplanarity of the set of spots). Now the spots were observed to cross paths.

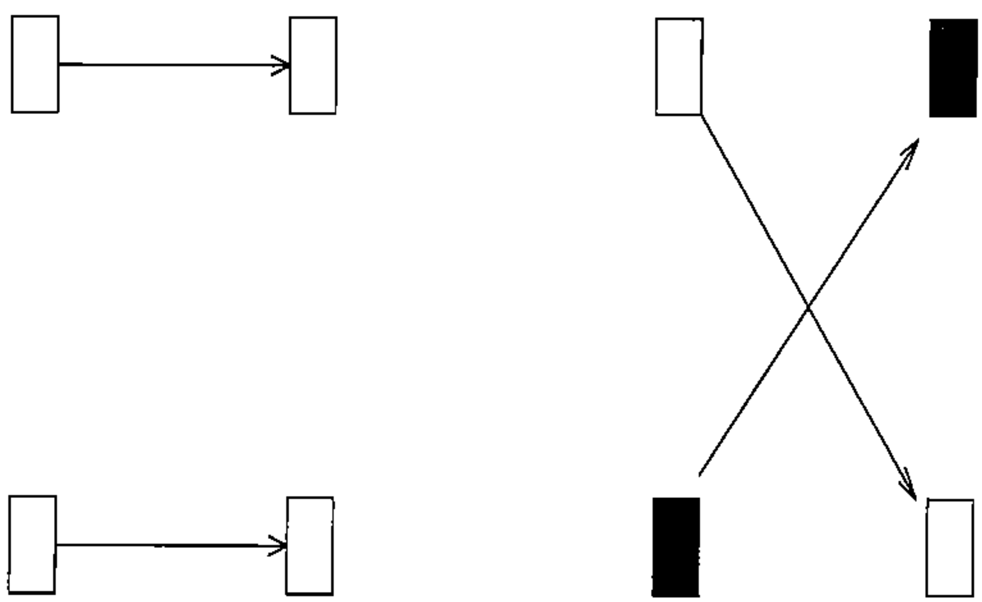

Figure 1: The stimuli given by Ramachandran & Anstis

Thus, motion correspondence was established on the basis of their ordinal depth properties (the targets having identical depths were corresponded) rather than on the basis of the 2D distances on the retina. This perceptual result is different from that of Ullman and Mutch *et al.* The difference is related to the fact that in Ullman's and Mutch *et al.*'s experiments different conditions were defined by changing the relative 3D distances among the targets, whereas in Ramachandran and Anstis's experiment different conditions were defined by changing the order in depth of the targets. The fact that the observer can reliably perceive order in depth and not 3D distances is consistent with the observed effect of depth in Ramachandran & Anstis's experiment, and the lack of any such effect in Ullman's and Mutch em et al.'s experiment.

It is, therefore, clear that Ullman's claim about the 2D motion correspondence requires revision in view of this conflicting experimental evidence. In fact, we believe that Ramachandran and Anstis's experiment is more likely to represent the actual role of depth in motion correspondence because it was designed in such a way that it did not preclude the reliable use of depth (as Ullman's and Mutch *et al.*'s experiments did). This new interpretation seems consistent with the anatomical results related to motion and form perception, which we discuss next.

Fig. 2 is a schematic of the motion pathway. The optic nerve carries the signal from the retinal ganglion cells to the Lateral Geniculate Nucleus. From here, the signal goes via area V1 (layers 4Cα, 4B) to area V2, which are parts of the striate cortex and extra striate cortex. Area V2(thick stripes) and layer 4B of area V1 project to the Middle Temporal(MT) lobe. Another input to MT comes from area V3, which in turn again receives input from layer 4B of area V1. Area MT sends projections via area MST(Medial Superior Temporal) to Area 7 in the Posterior Parietal cortex. It is well established that the Posterior Parietal cortex is involved in the representation of spatial locations. The damage to this area causes severe deficits in the ability to reach towards a target. Moreover, it is believed [2] that the motion processing that is used in solving correspondence are coded in the MT and MST areas. Consider now the neuoroanatomical stages of visual processing at which depth cues could be involved in solving correspondence. The thick stripes of area V2 have cells that are tuned to binocular disparity. The most consistent response from cells in this area is for stereoscopic depth [12]. Similar responses have also been obtained from some cells in layer 4B of area V1[3]. Thus, the depth information, i.e., distance or order in depth, is available to the area MT of the cortex. Since correspondence is probably established beyond area MT, in principle 3D positional data could be used by the correspondence process. Considering that such information is available, it might seem surprising that it would not be used. The only neurophysiological evidence that seems to be inconsistent with our analysis has been cited Dawson in [2]. According to him [2], Ullman's claim receives support from neurophysiological studies where it was shown that while there are neurons that code velocities in the frontoparallel plane, there are no neurons in area MT of the *Macaca fascicularis* that respond to motion in depth (Maunsell & Van Essen [15]). This evidence, however, bears closer scrutiny. Firstly, the best response from motion sensitive cells in MT was obtained if the frontoparallel motion was at a fixed disparity, i.e. a given distance in depth from the fixation point. Moreover, different units responded optimally to motion at different disparities. So even though there is no direct coding for motion in depth in area MT, there does seem to be coding of position in depth in this area, a point which the authors of [15] themselves make. Secondly, in area MST, which follows MT, there are cells which are sensitive to rotation in depth [26]. Thus existing anatomical and physiological evidence indicates that both position and motion in depth are coded in the visual system (areas MT & MST respectively). Such information is then likely to be available to Area 7, believed to be the site for establishing correspondence, which receives projections directly from MST and indirectly from MT.

It has to be pointed out that before binocular disparity of a given point can be used in motion correspondence, the visual system has to establish correspondence between the two 2D images coming from the two eyes (i.e. stereo correspondence). Evidently, the question of how this is done begs itself, since we claim that depth information is used to solve the motion correspondence problem. Stereo correspondence, however, is a special and restricted case of the more general correspondence that we are studying. It seems to be well established that stereo correspondence involves small horizontal disparities, and can be established by cooperative processes involving 2D information such as those described in [14]. In fact, humans can establish stereo correspondence even amongst random dot stereograms [9], where only retinal disparity can be used to solve correspondence (although, many 2D cues like shape, colour, size can improve the efficiency of the stereo correspondence). Stereo correspondence seems to be a part of the short range process in motion detection, unlike the more general case of correspondence, which involves the long range process. For a good review of the two process distinction, we refer the reader to an excellent paper by Petersik [18]. Thus, whereas stereo correspondence seems to be based on 2D cues only (distance, shape, colour, size), motion correspondence involves the depth information as well, derived from the binocular disparity. In fact, as shown by Ramachandran and Anstis, motion correspondence does not use many 2D cues, like shape or colour. This seems ecologically reasonable. Living organisms can change shape and colour in time, and therefore motion correspondence, which is used to perceive them, should not rely on these properties. However, at any given moment in time, an object has a particular shape and colour and therefore, stereo correspondence can, and should rely on these properties.

Consider now anatomical correlates of form perception and evidence for possible interaction between form and motion perception. Form perception involves areas in the IT region. Figure 3 shows a schematic of the early stages of the two pathways, form and motion. The analysis of the block diagram in Fig. 3 suggests that the perceived three dimensionality of a 2D line drawing in Ullman's experiments would come from the stages of the visual system in the Parvo pathway in the IT (InferoTemporal) areas. This information would then have to come back via feedback and cross links between the Magno and Parvo systems. The information has to come to the Magno system since this is where correspondence is established. However, Magno & Parvo are often viewed as separate processing streams, and the extent of their interconnection is very limited. It has even been claimed that there are no connections between them [12] whatsoever. This anatomical evidence is at the very least inconsistent with Ullman's interpretation of his experiment, which assumes that the object being perceived as three dimensional implies the availability of three dimensional information to the correspondence establishing centres in the brain. The alternative interpretation which we prefer is that when direct depth cues (from stereo and accommodation) are not present, like in Ullman's experiment or not reliable, like in Mutch *et al.*'s experiment, the human visual system establishes correspondence in two dimensions. To conclude from here that correspondence is always established in two dimensions is a non sequitur, and requires what in our opinion is a leap of faith. Ramachandran and Anstis's experiment used this aspect of depth which is perceived reliably and, as a result, depth was shown to be involved in motion correspondence. This psychophysical result appears to be consistent with the known facts about the processing streams in the visual cortex.

To summarise, we showed, by the reinterpretation of existing results, that studies seemingly supporting the claim that correspondence is solved in 2D only are certainly not conclusive to that end. They may be interpreted consistent with a correspondence in three dimensions paradigm, the interpretation we favour. In fact, this new interpretation seems more plausible in view of prior results on depth perception and Ramachandran and Anstis's study of motion correspondence at different disparities.

We also want to point out some interesting ramifications for computer vision. If the human visual system can use 3D correspondence on the basis of 2D images, it may be possible to improve existing algorithms for computer vision applications by studying its behaviour and using that information[8, 7]. As pointed out earlier in this paper, this may involve the need to use simplifying assumptions about the structure of the visual space (e.g., ordinal vs. metric). All existing computer vision approaches assume that the space is metric. However, weaker assumptions are likely to be more justified and lead to more efficient solutions [19, 20].

References

[1] H. Baker, *Depth from edge and intensity based stereo*, Ph.D. thesis, Computer Science Department, Stanford University, 1981.

[2] M.R.W. Dawson, *The how and why of what went where in apparent motion*, Psycho-

logical Review **98** (1991), 569–603.

[3] E.A. DeYoe and D.C. Van Essen, *Concurrent processing streams in monkey visual cortex*, Trends in Neurosciences **11** (1988), 219–226.

[4] Johnston E.B., *Systematic distortions of shape from stereopsis*, Vision Research **31** (1991), 1351–1360.

[5] A. Gruss, L.R. Carley, and T. Kanade, *Integrated sensor and range-finding analog signal processor*, IEEE Journal of Solid-State Circuits **26** (1991), no. 3, 184–191.

[6] R. Jain and I.K. Sethi, *Finding trajectories of feature points in a monocular image sequence*, IEEE Transactions on Pattern Analysis and Machine Intelligence **9** (1987), 56–73.

[7] A. Joshi and C.H. Lee, *On the problem of correspondence in range data and some inelastic uses for elastic nets*, Tech. Report CSD-TR-92-058, Department of Computer Science, Purdue University, 1992, submitted to IEEE Trans. Neural Networks.

[8] ———, *Using elastic nets for correspondence in range data*, Proceedings ICNN '93, San Fransisco, 1993.

[9] B. Julesz, *Foundations of Cyclopean Perception*, University of Chicago Press, Chicago, USA, 1971.

[10] Y.G. Leclerc and S.W. Zucker, *The local structure of image discontinuities in one dimension*, IEEE Transactions on Pattern Analysis and Machine Intelligence **9** (1987), 341–355.

[11] C.H. Lee and A. Joshi, *Correspondence problem in image sequence analysis*, Pattern Recognition **26** (1993), no. 1, 47–61.

[12] M. Livingstone and D.O. Hubel, *Segregation of form, color, movement and depth: anatomy, physiology and perception*, Science **240** (1988), 740–749.

[13] D. Marr, *Vision*, W.H. Freeman and Company, U.S.A, 1982.

[14] D. Marr and T. Poggio, *Cooperative computation of stereo disparity*, Science **194** (1976), 283–287.

[15] J.H.R. Maunsell and D.C. Van Essen, *Functional properties of neurons in middle temporal visual area of the macaque monkey, I & II*, J. Neurophysiol. **49** (1983), 1127–1167.

[16] S.P. McKee, D.M. Levi, and S.F. Bowne, *The imprecision of stereopsis*, Vision Research **30** (1990), 1763–1779.

[17] K. Mutch, I.M. Smith, and A. Yonas, *The effect of two dimensional and three dimensional distance on apparent motion*, Perception **12** (1983), 305–312.

[18] J.T. Petersik, *The two process distinction in apparent motion*, Psychological Bulletin **106** (1989), 107–127.

[19] Z. Pizlo, *Geometry of visual space*, Investigative Ophthalmology and Visual Science **34** (1993), 708, (Abstract).

[20] Z. Pizlo, R. Rosenfeld, and I. Weiss, *The geometry of visual space: about the incompatibility betweenscience and mathematics*, submitted to CVGIP, 1994.

[21] Z. Pizlo and M. Salach-Golyska, *Is vision metric?: comment on lappin and love*, Perception and Psychophysics **55** (1994), 230–234.

[22] V.S. Ramachandran and S.M. Anstis, *Perception of apparent motion*, Scientific American **254** (1986), 102–109.

[23] K. Rangarajan and M. Shah, *Estimating motion correspondence*, Proceedings Conference on Computer Vision and Pattern Recognition, 1991.

[24] M. Salach-Golyska, Z. Pizlo, S. Vyain, and J Standish, *3d shape perception*, Investigative Ophthalmology and Visual Science **34** (1993), 1082, (Abstract).

[25] V. Salari and I.K. Sethi, *Feature point correspondence in presence of occlusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990), 87–91.

[26] M.E. Sereno, *Neural Computation of Pattern motion*, The MIT Press, Cambridge, MA, USA, 1993.

[27] J.T. Todd and P. Bressan, *The perception of 3-dimensional affine structure from minimal apparent motion sequences*, Perception and Psychophysics **48** (1990), 419–430.

[28] J.T. Todd and J.F. Norman, *The visual perception of smoothly curved surfaces from minimal apparent motion sequences*, Perception and Psychophysics **50** (1991), 509–523.

[29] J.T. Todd and F.D. Reichel, *Ordinal structure in the visual perception and cognition of smoothly curved shapes*, Psychological Review **96** (1989), 643–657.

[30] S. Ullman, *Two dimensionality of correspondence process in apparent motion*, Perception **7** (1978), 683–693.

[31] ———, *The Interpretation of Visual Motion*, MIT Press, Cambridge, MA, USA, 1979.

[32] T.D. Williams, *Depth from camera motion in a real world scene*, IEEE Transactions on Pattern Analysis and Machine Intelligence **2** (1980), 511–516.

[33] D. Zhang, *Perspective invariant description of a planar point set and its application to matching*, Proceedings International Conference on Pattern Recognition, 1986.

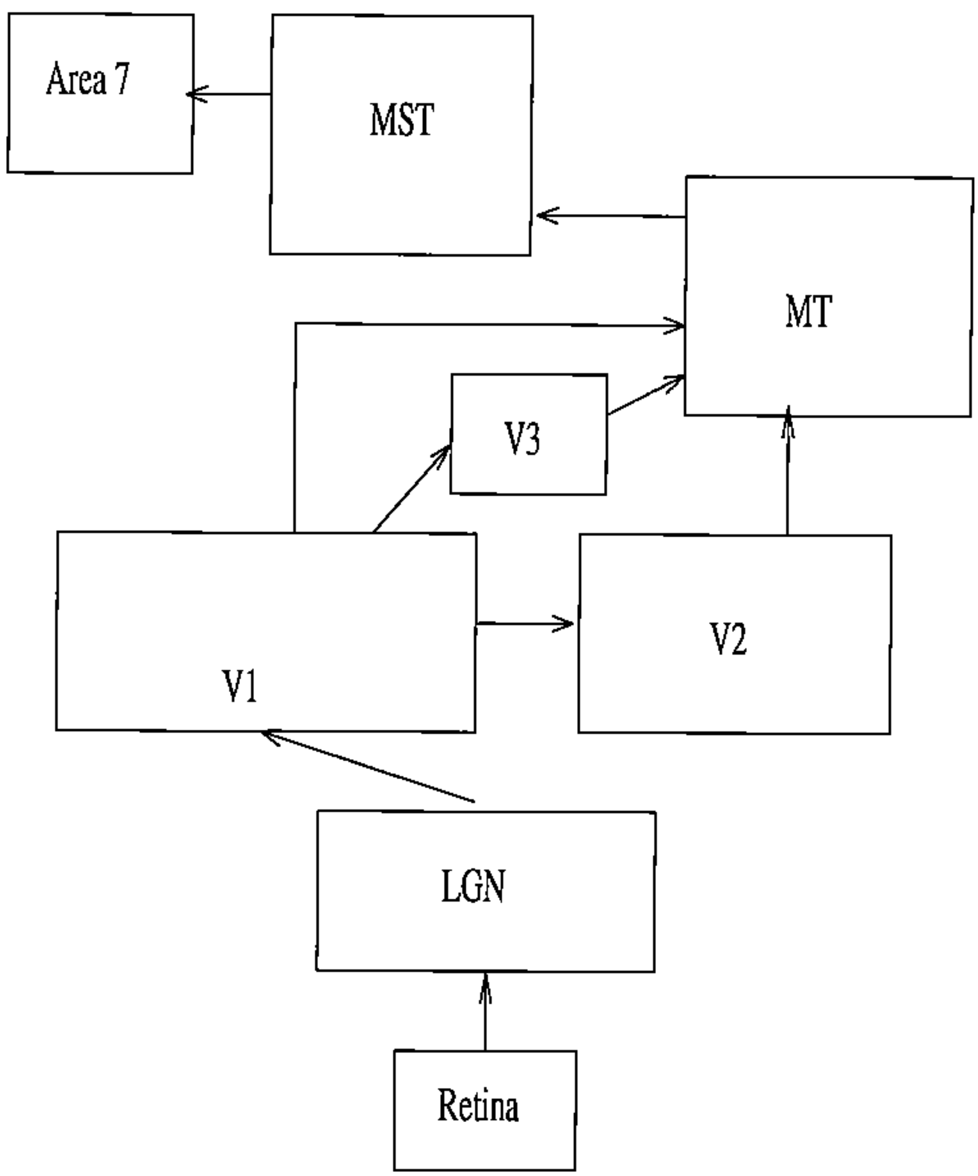

Figure 2: Pathway for Motion analysis in Visual Cortex

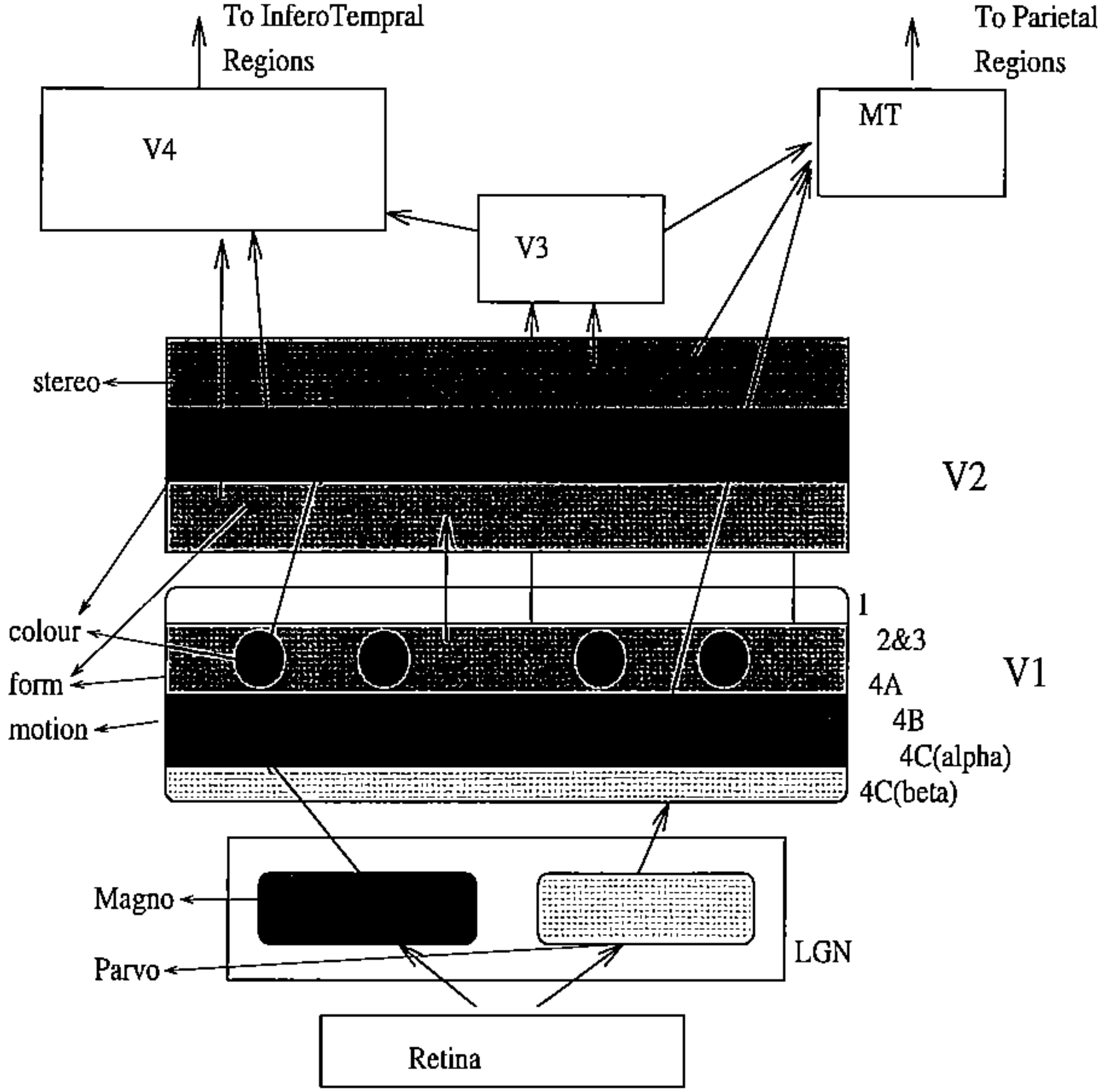

Figure 3: Schematic of Magno and Parvo Pathways in Early Vision